Nonmetric Sammon Mapping

Jan de Leeuw

November 19, 2025

TBD

Table of contents

Note: This is a working manuscript which will be expanded/updated frequently. All suggestions for improvement are welcome. All Rmd, tex, html, pdf, R, and C files are in the public domain. Attribution will be appreciated, but is not required. The files can be found at https://github.com/deleeuw/sammon

1 Sammon Loss

In engineering and computer science *Sammon mapping* is a popular multidimensional scaling (MDS) method. The technique was introduced in Sammon Jr (1969). It was originally intended to map points from a higher-dimensional Euclidean space into points in a lower-dimensional Euclidean space by approximating the given higher dimensional distances by best-fitting lower dimensional ones. The MASS package for R implemented the sammon() function, which generalizes the original idea by allowing the higher-dimensional distances to be replaced by any positive symmetric matrix of dissimilarities. This was again generalized in the packages stops (Rusch, Mair, and Hornik (2023)) and smacofx (Rusch et al. (2025)) where optimization is over low-dimensional configations and over power transforms of the dissimilarities.

The Sammon loss function is

$$\sigma(X,\Delta)=\frac{1}{\sum\sum_{1\leq i<j\leq n}w_{ij}\delta_{ij}}\mathop{\sum\sum}_{1\leq i<j\leq n}w_{ij}\frac{(\delta_{ij}-d_{ij}(X))^2}{\delta_{ij}}. \tag{1}$$

Without loss of generality we assume the weights w_{ij} add up to one.

In metric (ratio) MDS we minimize over X for fixed Δ. This is a standard metric scaling problem with weights $w_{ij}\delta_{ij}^{-1}$. We use majorization steps to decrease loss. This is identical to the ratio sammon option in smacofx. In the non-metric case we use alternating least squares, i.e. we alternate majorization steps with minimizing over Δ, satisfying the ordinal constraints, for fixed X. As far as I know this has not been implemented before, and is somewhat non-standard. First, use the fact that $\sigma(\lambda X,\lambda\Delta)=\sigma(X,\Delta)$

$$\min_X\sigma(X,\Delta)=\min_X\sigma(X,\lambda\Delta), \tag{2}$$

i.e. the problem is homogeneous of degree zero in Δ. Thus we can require without loss of generality that

$$\mathop{\sum\sum}_{1\leq i<j\leq n}w_{ij}\delta_{ij}=1. \tag{3}$$

With this normalization we have

$$\sigma(X,\Delta)=1-2\mathop{\sum\sum}_{1\leq i<j\leq n}w_{ij}d_{ij}(X)+\mathop{\sum\sum}_{1\leq i<j\leq n}w_{ij}\frac{d_{ij}^2(X)}{\delta_{ij}} \tag{4}$$

Thus minimizing ω for fixed X over non-negative, isotone and normalized Δ means minimizing the third term on the right. And this means minimizing a separable, differentiable, and strictly convex function over a polyhedral convex set.

2 PAVA for Sammon

Define the extended real valued function f equal to

$$f(x) := \sum_{i=1}^{n} w_i \frac{y_i}{x_i} \tag{5}$$

if $x_i > 0$ for all i, and to $+\infty$ otherwise. Thus the effective domain $\mathrm{dom}(f)$ is the interior of the positive orthant.

Problem $\mathfrak{P}$ is defined as minimization of f over the intersection of the polyhedral convex cone $\mathcal{K}$ defined by $0 < x_1 \leq \cdots \leq x_n$ and the affine set $\mathcal{A}$ defined by $w'x = 1$. We assume that all w_i and all y_i are positive, and that the w_i add up to one. In the context of Sammon mapping the y_i are squared distances and the x_i are the transformed dissimilarities.

The first and second partials of f are

$$\mathcal{D}_i f(x) = -w_i \frac{y_i}{x_i^2}, \tag{6}$$

$$\mathcal{D}_{ii} f(x) = 2w_i \frac{y_i}{x_i^3}, \tag{7}$$

and $\mathcal{D}_{ik} f(x) = 0$ if $i \neq k$. This shows f is convex and twice differentiable on $\mathrm{dom}(f)$, and thus on $\mathcal{K} \cap \mathcal{A}$.

First some preliminary results.

Lemma 2.1.

$$\min_{x \in \mathcal{K} \cap \mathcal{A}} f(x) \leq w'y \tag{8}$$

Proof. The vector x with $x_i = 1$ for all i is in $\mathcal{K} \cap \mathcal{A}$. □

It follows from Lemma 2.1 that we can add the constraint $f(x) \leq w'y$ to the minimization problem and still have the same minimum and minimizer.

Lemma 2.2. *Suppose $y_1 \leq \cdots \leq y_n$. Then*

$$\overline{x}_i = \frac{\sqrt{y_i}}{\sum_{k=1}^{n} w_k \sqrt{y_k}}. \tag{9}$$

Proof. The necessary conditions for a minimum of f on $\mathcal{A}$ are

$$-w_i\frac{y_i}{x_i^2} = \lambda w_i, \tag{10}$$

for all i, together with the side condition $w'x = 1$. From (10) it follows that the solution $\overline{x}$ must be proportional to $\sqrt{y}$, where the square root can have either sign. Of the 2^n solutions only one is in the effective domain of f, the one for which all all square roots are taken with a positive sign. This solution is also in $\mathcal{K}$. Applying the side condition gives (9). □

The next rather trivial lemma deals with the case $n = 1$, in which w, x and y are one-element vectors, which identify with the corresponding scalars.

Lemma 2.3. *If $n = 1$ then the minimizer $\overline{x}$ is equal to one and the minimum is y.*

Proof. w adds up to one, so $w = 1$. Also wx must be one, so $\overline{x} = 1$. □

The following theorem is of prime importance, because it shows problem $\mathfrak{P}$ can be solved with a with a variation of the Pool Adjacent Violaters Algorithm (PAVA). For the details on PAVA, see for example De Leeuw, Hornik, and Mair (2009).

Theorem 2.1. *Suppose $\overline{x}$ is the optimum solution. If $y_i \geq y_{i+1}$ then $\overline{x}_i = \overline{x}_{i+1}$.*

Proof. We show that $y_i \geq y_{i+1}$ and $\overline{x}_i < \overline{x}_{i+1}$ leads to a contradiction. A necessary and sufficient condition for $\overline{x}$ to be the optimum solution is

$$(x-\overline{x})'\mathcal{D}f(\overline{x}) = -\sum_{i=1}^n w_i\frac{y_i}{\overline{x}_i^2}(x_i - \overline{x}_i) = f(\overline{x}) - \sum_{i=1}^n w_i\frac{y_i}{\overline{x}_i^2}x_i \geq 0 \tag{11}$$

for all $x \in \mathcal{K} \cap \mathcal{A}$ (Hiriart-Urruty and Lemaréchal (1993), Theorem 1.1.1, page 293).

Now suppose $\overline{x}_i < \overline{x}_{i+1}$. Then for $\epsilon > 0$ small enough

$$z = \overline{x} + \epsilon\left(\frac{e_i}{w_i} - \frac{e_{i+1}}{w_{i+1}}\right) \tag{12}$$

is also in $\mathcal{K} \cap \mathcal{A}$. In (12) we add a small amount to $\overline{x}_i$ and subtract a small amount from $\overline{x}_{i+1}$, while leaving all other elements of $\overline{x}$ unperturbed. Since

$$\sum_{i=1}^n w_i\frac{y_i}{\overline{x}_i^2}z_i = f(\overline{x}) + \epsilon\left(\frac{y_i}{\overline{x}_i^2} - \frac{y_{i+1}}{\overline{x}_{i+1}^2}\right) \tag{13}$$

we must have

$$\frac{y_i}{\overline{x}_i^2} \leq \frac{y_{i+1}}{\overline{x}_{i+1}^2} \tag{14}$$

But $y_i \geq y_{i+1}$ and $\overline{x}_i < \overline{x}_{i+1}$ implies

$$\frac{y_i}{\overline{x}_i^2} > \frac{y_{i+1}}{\overline{x}_{i+1}^2} \tag{15}$$

and thus $\overline{x}$ cannot be the optimal solution. □

Theorem 2.1 can be used to replace problem $\mathfrak{P}$ of size n with a problem $\tilde{\mathfrak{P}}$of the same type, but of size $n-1$. If $y_i \geq y_{i+1}$ we remove these two y-values and put the single value

$$\tilde{y}_i := \frac{w_i y_i + w_{i+1} y_{i+1}}{w_i + w_{i+1}}$$

in their place. This new value gets the weight $\tilde{w}_i := w_i + w_{i+1}$.

Theorem 2.2. *Suppose $\hat{x}$ solves the monotone regression problem for the weighted least squares norm*

$$\hat{x} := \underset{x_1 \leq \cdots \leq x_n}{\mathit{argmin}}\ (x-y)'W(x-y)$$

then

$$\overline{x}_i = \frac{\sqrt{\hat{x}_i}}{\sum_{k=1}^n w_k \sqrt{\hat{x}_k}}$$

solves problem $\mathfrak{P}$.

Proof. We use Theorem 2.1 repeatedly until we have a problem $\mathfrak{P}$ of size r for which $\tilde{y}_1 < \cdots < \tilde{y}_r$. We can use the same sequence of pooling adjacent violators as in least squares monotone regression, and we find the same weighted average pooled values in each step. When we have reduced the problem to a strictly increasing $\tilde{y}$ sequence we apply Lemma 2.2 (and if $r = 1$ we apply Lemma 2.3). We then expand each block again to a length equal to the number of averaged elements. □

A small example illustrates the PAVA variationb. Start with y equal to $(1,2,3,1,2,3,5,1)$ and w equal to $(1,1,1,1,1,1,1,1)$. Merge elements 3 and 4 to get $y = (1,2,2,2,3,5,1)$ and $w = (1,1,2,1,1,1,1)$. Now merge elements 2, 3, and 4 to get $y = (1,2,3,5,1)$ and $w = (1,4,1,1,1)$. Merge 4 and 5 to get $y = (1,2,3,3)$ and $w = (1,4,1,2)$ and finally merge 3 and 4 to get $y = (1,2,3)$ and $w = (1,4,3)$. Thus

$$\overline{x} = \frac{1}{1+4\sqrt{2}+3\sqrt{3}}(1,\sqrt{2},\sqrt{2},\sqrt{2},\sqrt{2},\sqrt{3},\sqrt{3},\sqrt{3})$$

References

De Leeuw, Jan, Kurt Hornik, and Patrick Mair. 2009. "Isotone Optimization in R: Pool-Adjacent-Violators Algorithm (PAVA) and Active Set Methods." *Journal of Statistical Software* 32 (5): 1–24.

Hiriart-Urruty, Jean-Baptiste, and Claude Lemaréchal. 1993. *Convex Analysiis and Minimization Algorithms Volume 1: Fundamentals*. Springer.

Rusch, Thomas, Jan de Leeuw, Lisha Chen, and Patrick Mair. 2025. *Smacofx: Flexible Multidimensional Scaling and 'Smacof' Extensions*. https://doi.org/10.32614/CRAN.package.smacofx.

Rusch, Thomas, Patrick Mair, and Kurt Hornik. 2023. "Structure-Based Hyperparameter Selection with Bayesian Optimization in Multidimensional Scaling." *Statistics and Computing* 33 (28): 1–18.

Sammon Jr, John W. 1969. "A Nonlinear Mapping for Data Structure Analysis." *IEEE Transacttions on Computers* C-18 (5).